W.J. Usery Workplace Research Group Paper Series

Working Paper 2010-2-2
February 2010

Institutional Economics and the Theory of What Unions Do

Bruce E. Kaufman
Georgia State University

Georgia State University | Andrew Young School of Policy Studies

Institutional Economics and the Theory of What Unions Do

Bruce E. Kaufman

Professor, Department of Economics, Georgia State University, Atlanta GA, USA

Research Fellow, Griffith Business School, Griffith University, Brisbane QLD, AU

February 23, 2010

Paper submitted for "Institutions, Work, and Employment Relations" Symposium

Institutional Economics and the Theory of What Unions Do

Abstract: The theory of trade unions is re-examined using principles and ideas of institutional economics. An institutional perspective provides a more balanced and inclusive portrait of what unions do; it also demonstrates flaws and biases in the standard neoclassical account that lead to overly negative conclusions. Unions can either be "monopsony-reducing" or "monopoly-creating" in their economic and governance functions and may thus in some situations improve economic performance and welfare but it others harm them. The "optimal" level of union density depends on the breadth and depth of market and governance failures and an assessment of the feasibility, benefits and costs of alternative institutional solutions to labor problems.

This paper uses ideas and principles of institutional economics (IE) to reexamine a century old issue: the economic function and effect of trade unions. The end-product is distinctive not only because it casts the IE perspective into a more analytical and price theoretic framework but also for the new and revisionist perspective it yields on the economic performance and welfare effects of unions.

The roots of labor economics in the USA go back to the work of the original institutional economists in the late 19th/early 20th centuries, such as Ely, Commons, Hoxie and Slichter, and the topic that originally drew them to the study of labor was trade unionism (e.g., Ely 1886; Commons 1905). Also important were European writers, particularly the Webbs' magisterial *Industrial Democracy* (1897). With the advent of the Chicago School and resurgence of neoclassical economics in the early post-World War II period, the Wisconsin version of institutionalism gradually disappeared from the mainline literature of labor economics and took up an increasingly shadowy existence in the nearby field of industrial relations (Kaufman 2006). As it did, the style of analysis and tenor of debate on trade unions firmly shifted from an historical, participant-observer, multidisciplinary and open-minded-to-friendly approach to a price theory, uni-dsciplinary, econometric, and skeptical-to-critical approach.

The publication of Freeman and Medoff's *What Do Unions Do?* (1984) book in the mid-1980s didn't change the microeconomic/econometric type of analysis but did endeavor with some new theoretical ideas and empirical evidence to shift the verdict on unionism toward a more neutral or even positive direction. A fair assessment seems to be threefold: first, that their revised theory (the "exit/voice" or "two faces" model) made a lasting but not transformative impact on the theory-side of the literature; second, their revised empirical assessment of union effects had some selective but less widely accepted success in moving the verdict of economists toward a more neutral/less negative position on union welfare effects; and, third, they mostly failed to dent the core "labor monopoly" model that economists carry in their heads and use in research and textbooks when it comes to describing the discipline's base-line view of what unions do (Bennett and Kaufman 2007).

Research interest in the USA on unions has noticeably declined during the last decade and the topic now has a distinctly marginalized-to-almost dead feeling, owing no doubt to the parallel steep decline in unions themselves. In 2009, for example, private sector density fell to 7% and a majority of members were located in the public sector. The picture is less dire for unions in other countries but the trend is nonetheless downward in most.

Meanwhile, and largely outside the purview of labor economics, from the 1980s onward institutional economics enjoyed a rebirth in the USA and Europe, albeit in the guise of a reconfigured "new" institutional economics (NIE) with its primary intellectual roots traced not to Commons but Coase. The NIE was originally positioned by proponents (e.g., Williamson 1985) as mostly a new line of thought vis-à-vis the original institutional economics (OIE) and more of an extension of the neoclassical tradition than an alternative, although in recent years some shift has occurred in both areas. The NIE largely focuses on industrial-organization topics and until

the last decade made only modest contact with labor economics and very little with unions (see Dow 1997). Here too, however, a shift is discernible with considerable more interest among labor economists in the economic analysis of contracts, internal firm organization, and the performance effect of institutions and, further, one recent book applies transaction cost theory to trade unions (Barker 1997).

Also of note, institutionalism in Europe has enjoyed a similar resurgence. The economics branch draws eclectically from OIE (including German/English social/historical economics) and NIE while a complementary branch in sociology and political science – most visible in the "varieties of capitalism literature" (Hall and Sokice 2001) -- draws its intellectual inspiration from writers such as Polanyi, Durkheim and Weber. The sociology branch, in particular, pays considerable attention to labor market and employment issues and operates from a theoretical base that generates a more sympathetic perspective on institutional regulation/coordination of labor markets and the economic role of trade unions (Crouch 1993; Streeck 2001; Amable 2003).

The contribution of this paper is to take these disparate strands of institutionalism and weave them together for purposes of examining the economic role and effect of trade unions. Primary emphasis is given to OIE and NIE. The OIE perspective on trade unions has been described in detail elsewhere (Kaufman 2007a and citations therein) so I introduce this material here only where necessary for context and documentation. Instead, the major goal of this paper is to take the institutional perspective, develop it in a more analytical way, and apply it to the analysis of trade unions. Purposes of both "destruction" and "construction" are served. The destruction part is to show that the standard neoclassical analysis of unions is in certain respects fundamentally flawed and biased; the construction part is to develop an alternative IE-based analysis of unions. In the realm of positive economics, new concepts, models and hypotheses are

advanced. In the realm of welfare (normative) economics, an IE perspective shifts the discipline of economics from a base-line and instinctively negative "guilty until proven innocent" verdict on unions to a more impartial and balanced perspective in which theory suggests the verdict can go either way and the ultimate decision is based on the weight of empirical evidence pro and con.

The Neoclassical Analysis of Unions

The neoclassical analysis of unions is well-known and developed. Detailed reviews are provided in Oswald (1982), Farber (1986), Hirsch and Addison (1986), Pencavel (1991), Booth (1995), Naylor (2003), Metcalf (2005) and Kaufman (2002a, 2007b). I focus here on the base-line model in this literature, widely known as the monopoly union model, with certain extensions and qualifications introduced later.

The monopoly "on the demand curve" (or "right-to-manage") model of unions is shown in *panel a* of Figure 1. The labor market absent trade unionism is competitive. The wage rate and level of employment, W_1 and L_1 (point A), are determined by the labor demand (D) and supply curves (S). Assume a union now organizes the firms in this market. The union has a preference function defined in terms of higher wages and larger employment, giving rise to the union indifference curve I_1. The tangency point between I_1 and the demand curve D gives the union's preferred wage (point B). The simplest model assumes the union has the bargaining power to set the wage, so it chooses W_2; given this, employers then chooses the level of employment at L_2. Because the union has raised the cost of labor, and absent any countervailing economies, employment in the market contracts through layoff of L_1-L_2 workers.

[Insert Figure 1 here]

Seen from this vantage point a union is similar to a labor market monopoly; that is, it uses its monopoly of the supply of labor and threat to strike to coerce firms to pay an above-competitive wage for labor. Like a monopoly in the product market, a union leads to certain predictable economic effects. A union transfers wealth and workplace control from employers to the union and its members; the wealth transfer, in turn, comes partly from firm owners (lower short-run profit) but also consumers (higher product prices) and non-union workers (a lower wage in the non-union sector as laid-off union workers add to the competition for jobs).

Another but perhaps less obvious loser in this wealth redistribution is society at large. The reason is that a union wage increase distorts relative factor prices and leads to a misallocation of resources and welfare loss. One demonstration of this point is given in *panel a* by the triangle ABC (shaded area). It depicts the deadweight welfare loss measured in terms of reduced consumer and producer surplus. A second way to illustrate the union harm to efficiency and welfare is in *panel b*. By the neoclassical first welfare theorem, the competitive outcome is located somewhere on the economy's production possibility frontier, such as point A. Like a product market monopoly, a union distorts relative prices, leads to a resource misallocation, and moves society inside the frontier to a place such as point B. The distance A-B represents the decrease in production of goods X and Y (e.g., guns and butter) and parallel loss in jobs that arises from unionism in the labor market.

The monopoly model is simplistic in a number of respects; nonetheless, many-to-most economists take its predictions as a reasonable first approximation of what really happens under collective bargaining. In this spirit, Hirsch and Addison (1986) state, "The monopolistic view of unionism, firmly held by most economists, begins with the presumption that unions raise wage rates above competitive levels in the union sector" (p. 21). They go on to say, "Society suffers

net welfare losses from unionism owing to the resulting inefficient factor mix and the misallocation of resources between the union and non-unionn sectors." (p. 22). Similarly, Booth (1995) remarks, "The standard view of trade unions is that they are organizations whose purpose is to improve the material welfare of members, principally by raising wages above the competitive level. There is little dispute that unions are frequently able to push wages above the competitive level – what is called the 'monopoly' role of trade unions …. and introduce into the economy a variety of distortions and inefficiencies" (p. 51).

The monopoly model has been amended, challenged and revised in various ways. For purposes of this discussion, three extensions merit mention: the median voter, exit-voice, and efficient contract models. The first two help move the theoretical analysis of unions in an institutionally compatible direction, the third helps bring to light a fundamental flaw in the standard monopoly model.

The median voter model modifies the union's objective function. Following Dunlop's original formulation (Dunlop 1944), the standard model treats a union as akin to a firm and then specifies a union objective function that parallels the profit maximization function of the firm. Numerous versions have been suggested, such as wage bill, rent and expected utility maximization. As Dunlop recognized (but proceeded to ignore), and as also noted by a long line of later writers, the "union as a firm" analogy breaks down at several crucial points (Ross 1948; Martin 1980; Barker 1997). One is that a firm owns the products it sells and keeps the surplus between revenue and cost; a union, however, does not own labor (the members do) and does not collect any surplus of wages or rents. A second is that the nature of property rights in a firm creates a unanimity of interest among shareholders in the goal of profit maximization (since all share in the proceeds) while in most unions no such unanimity exists. The reason is that most

unions assign differential property rights (i.e., internal rules that determine the distribution of individual benefits and costs) to members based on characteristics such as length of membership, job seniority, occupation, and geographical location (and, in an earlier era, race and gender) and these differential property rights create a corresponding diversity of membership interests. Thus, all shareholders want the firm to strive for maximum profit but union members are likely to differ on what they consider is the union's optimal bargaining objective (e.g., a high, medium or low wage demand).

The challenge facing economists is to derive a tractable and logically consistent union objective function from the diverse preferences of the membership. One approach is to take the "as if" modeling strategy and simply specify a plausible union objective function, such as in *panel a*, and see if it generates hypotheses and predictions that accord well with the empirical evidence.

A second and more IE-oriented approach is to make assumptions about the internal property rights in the union that allow aggregation of a consistent union objective function. From an IE perspective -- given its methodological preference for empirically congruent assumptions -- the least satisfactory but most analytically tractable tact is to greatly simplify the aggregation problem by essentially assuming it away, say by positing union employment is determined by random draw (Oswald 1982; Naylor 2003). More preferable is to distinguish alternative regimes of property rights, such as the dichotomy between craft and industrial unions (Kerr, 1954; Martin, 1980). The former, for example, typically rations jobs by rationing membership cards and giving preference to local area members; the latter, on the other hand, rations jobs principally through seniority provisions. Martin (1980) shows that a union's wage policy differs substantially between the two types of property rights regimes. Yet another approach involves

the median voter model, imported from the public finance literature of economics. This model is utilized below so discussion of its properties is deferred to later.

Institutional Analysis of Unions: Background Concepts and Ideas

This section of the paper provides an overview of concepts, ideas and principles relevant to an institutional analysis of unions. It is divided into three parts. The first very briefly outlines key concepts and principles in institutional economic theory; the second provides a modestly more detailed outline of the IE perspective on labor markets; and the third does the same for the IE perspective on trade unions. The subsequent section then uses these concepts and ideas to develop a more analytical and IE-oriented analysis of unions.

IE Concepts and Principles

Since many economists are unfamiliar with the institutional paradigm – admittedly to large degree because institutional economists have failed to spell it out very well, the place to start is briefly setting out key concepts and ideas. They include:

Ownership and property rights: ownership and property rights are the foundation of IE theory, for they determine the institutional structure of an economy and how it performs; without prior specification of property rights and ownership, fundamental economic constructs such as commodities, production functions, demand and supply curves, and efficient allocations have no basis.

Institutions: institutions are bodies of rules, both formal and informal and explicit and tacit, that are built out of property rights (broadly defined) and define the rules of the economic game and the

resources, constraints, opportunity sets, incentives, and strategic interdependencies faced by economic agents. Institutions determine the structure (and existence) of markets which, in turn, determine their behavior and performance. All economic behavior takes place within institutions.

Institutions as Governance Structures: institutions exist to structure and coordinate production and exchange and they do so by governing the terms, conditions and processes underlying contract negotiation, fulfillment and enforcement. Firms are not merely production functions and markets are not an abstract meeting place of demand and supply; rather, they and all other institutions are political constructs that distribute power, define relationships, give access to resources, and determine individual benefits, and costs.

Sovereignty: the political construction of institutions implies economics is inescapably "political economy" and hence politics and economic outcomes are inextricably linked. Central to political economy, therefore, is the concept of sovereignty since it determines who sets the rules of the game and whose preferences count in this process. Economic actors individually and collectively, therefore, seek to capture and use the power of sovereignty to shape the institutions to promote their interests and ethical viewpoints.

Behavioral/social model of the human agent: people are modeled as largely purposeful and self-interested, but decision-making is subject to bounded rationality and behavior is influenced by social interdependencies, emotions, and ethical precepts. Workers are an expressly human factor of production.

Transactions and transaction cost: a transaction is a legal transfer of ownership; transaction cost is the real resources used to effectuate and enforce this transfer. Positive transaction cost is the rule.

Modes of coordination: economies have alternative institutional modes for coordinating transactions, including markets and organizations. Markets use prices as the coordinating device;

organizations use command and administration. Because of positive transaction cost, market exchange often involves an element of bargaining.

Power: power is the ability to satisfy one's desires and obtain a greater share of an institution's scarce goods (material and non-material).

Reasonable value: economic agents individually and collectively have a notion of what is fair and reasonable; whenever an outcome or process falls outside the bounds of reasonableness, they undertake action to redress the imbalance. Possible actions include restricting supply, withdrawing from the relationship, and seeking a more equitable matrix of rules and rights.

Evolution: the interaction between outcomes and institutional structure causes economies to evolve over time in a process of cumulative causation along different path-dependent trajectories.

IE Perspective on Labor Markets

A thumbnail sketch of the institutional perspective on the typical non-union labor market is given by Lloyd Reynolds (1954), based on extensive case study and participant-observer evidence. He states (p. 543) as a general principle, "[I]t is apparent that local labor markets in this country …. are not highly competitive." He goes on to amplify on this observation (p. 549):

> "Only in theory, then, does the 'competitive labor market' provide an alternative to wage determination through collective bargaining. The practical alternative is collective bargaining *versus* wage-setting by employers with rather weak competitive checks. Under non-union conditions, the immobility of the majority of workers plus the unsystematic selection of jobs by those in search of work gives employers wide latitude in determining wage rates and other conditions of employment. An employer can offer terms considerably below those generally prevailing in the area and still secure an adequate labor force. He is subject to serious competitive pressure mainly at the peak of business cycles, when job opportunities in other plants are relatively plentiful. Even after years of high unemployment, one still finds large differences in the wages offered by different employers for the same jobs."

Thirty years later Reynolds (1988: 138) allowed that he may have over-emphasized the seriousness of imperfect information and impediments to labor mobility and, further, his observations were heavily shaped by studies of manufacturing plants and workers. Nonetheless, in broad outline Reynolds captures the reality of non-union labor markets as seen by most institutionally-oriented labor economists. John Dunlop, for example, stated in 2002 his view that institutions "do not come into the picture and distort some 'perfect' wage structure, because there is no such thing. In the real world there are all kinds of distortions and inequities built into the wage structure" (quoted in Kaufman, 2002b: 338). That the institutional view of non-union labor markets is broadly accurate is also suggested by two in-depth empirical reviews by Richard Freeman (1984; 1988); in the former he compares Reynolds' case study conclusions against modern econometric evidence and concludes that Reynolds' work was "on target in its picture of the labor market" (Freeman 1984: 219).

Synthesizing from a large literature, the IE view of labor markets may be summarized with these key points:

Labor a Human Factor: labor is embodied in human beings and therefore the labor input comes to the market place and production process with a full range of cognitive limitations, multiple motivations, and emotions. Labor exhibits purposeful behavior organized around a core of self interest but constrained and supplemented by bounded rationality, interdependent preferences, social norms, ethical precepts, and varying degrees of non-logical sentiments. For modeling purposes, labor is for many purposes distinct and unique relative to other factors and goods.

The Employment Relationship: Workers most often provide labor services in an institution called the employment relationship (ER). Unlike competitive goods markets, the

buyer and seller of labor are embedded in a personal, long term, and socially interdependent relationship where many things besides price mediate and coordinate the exchange. It is also essential to recognize that the employment contract is a rental agreement for labor services.

Incomplete Employment Contracts: Because of bounded rationality, imperfect and asymmetric information, fundamental uncertainty, and the interdependent and complex nature of production tasks, transaction cost is both positive and large in most ERs. As a result, employment contracts are necessarily incomplete, contingent and open-ended and subject to numerous forms of externality, public good, moral hazard, opportunism, principal-agent problem, and tacit bargaining.

Imperfect Competition: labor markets are by their nature not only imperfect (in the economist's sense) but among the most imperfect in the economy. Competitive forces are certainly present and the demand/supply model has some degree of explanatory power; nonetheless, in the short-medium run most labor markets exhibit substantial wage rigidity, constraints on labor mobility, and in most years excess labor supply (involuntary unemployment). These conditions mean labor markets are not self-regulating and are partially coordinated by other means (e.g., labor quantity and quality adjustments); they do not necessarily mean that workers are exploited (although some are) and efficiency suffers (although it may). Indeed, labor markets are imperfect in part because restraints on competition promote the interests of both firms and workers and lead to efficiency gains in production.

Segmented Markets and Alternative Employment Systems: Although most labor markets are imperfect in a substantively important sense, they differ greatly in terms of structure, conduct and performance. The aggregate labor market is divided into segments more complex and variegated than the standard competitive versus monopsony categories (even with

monopsony defined broadly). Segmentation arises from factors that impede competitive forces and the flow of labor across firms and markets, including institutional rules (e.g., seniority systems, occupational licensing), different educational requirements, firm-specific skills, discrimination and social norms, and job search costs and the human desire for security. Differences in production systems and employment systems within firms also influences labor market structure/segmentation. Firms can be arrayed on a continuum from "advanced" and "high road" firms at the top using a human capital and high involvement employment system where workers are treated as a valuable human resource and semi-fixed cost to "backward" and "low road" firms at the bottom end using a commodity and low involvement employment system where workers are treated as a variable cost hired hand. Firms choose different human resource management (HRM) strategies and employment systems both within market segments and across segments, giving rise to considerable dispersion in terms and conditions of employment. Some elements of this dispersion are "equalizing," many others, however, are non-equalizing.

Firms and Internal Labor Markets: Employment systems are closely related to the concept of an internal labor market (ILM). Firms are more than a production function (or technical system); rather, they are also hierarchial organizations with a vertical command and control structure, management-created and administered HRM systems and an internal social system ("work society") made of a network of social/status relations and expectations. The structure of the production process and rules and procedures of the HRM system create internal labor markets (ILMs) in which management coordination and HRM practices partially supplant external market forces as the determinant of wages, employment conditions, and mobility patterns. ILMs also provide an important vehicle for in-house workforce training and development; likewise, by creating vertical lines of promotion they also foster greater employee

motivation, loyalty, and job tenure. All firms are "organized" in the sense workers and managers interact in a web of social relations that help govern and coordinate the production process, including work pace, task performance, and cooperation with management. The strength, structure and formalization of ILMs, HRM systems and workplace social systems vary considerably across labor market segments; they are most developed and influential in advanced "primary" sector firms and least developed and most porous in small/low end "secondary" firms.

Inequality of Bargaining Power: employers both individually and as a group have a power advantage over individual workers in both external wage bargaining and internal firm governance due to their legal authority over work (the "master-servant relationship"), control of the supply of jobs, the perishability of labor services (inability to inventory), workers' limited hold-out ability (from limited financial reserves, significant fixed costs of family subsistence), costly job search and restricted job opportunities, and tilted legal rules and resource endowments. These conditions create a "tipped playing field" both "within the market" and "before the market" that favor firms' interests in exchange and governance relationships, thus allowing employers to capture a disproportionate and possibly unjust/unreasonable share of economic surplus, workplace control, and life satisfaction.

Looking within labor markets and firm governance systems, some groups of workers with high education, valuable skills, jobs in primary sector firms or employers using a high road HRM strategy suffer little bargaining disadvantage and, indeed, may reap large rents due to inelastic supply or restricted competition. Other workers, however, experience this inequality of power in the form of lower pay, fewer benefits, worse treatment, or arbitrary termination than would exist if competition was truly balanced. Most likely among this group are people in vulnerable or discriminated-against demographic groups, workers with little education or few

marketable skills, and employees in secondary or low road firms. Workers' inequality of bargaining power (IBP) diminishes for all groups (or turns into an advantage for some) as the aggregate unemployment rate falls and employers aggressively compete for labor; it worsens, on the other hand, during periods of rising unemployment and lack of jobs and affects nearly all work groups in serious recessions or a depression.

Workers may suffer IBP "before the market" if the structure of legal rules gives a distinct bargaining advantage to firms (e.g., open immigration, no unemployment insurance, employment-at-will) or if firms are giant corporations with deep pockets and families have only one bread-winner with several dependents and meager savings. In this case, the labor market may well be "competitive" in the economist's sense yet the workers' supply curve lies sufficiently far to the right (because of a low reservation wage) that the competitive wage (and other terms and conditions of employment) yields a very unequal distribution of income and small-to-meager economic penalties (e.g., compensating wage differentials) for unsafe working conditions and harsh treatment. The fact that labor is rented, rather than owned like capital, provides yet additional incentives for employers to skimp on labor maintenance.

Cooperation, Trust, Fairness and Job Security: Production is in most cases an interdependent process that requires active cooperation among workers and managers. The degree of cooperation (including work effort) is a choice variable for workers; low cooperation typically means low productivity and profits and high cooperation means the reverse. Many factors influence workers' willingness to cooperate but among the most important are trust, fairness and job security.

The employment relationship has significant elements of a Prisoner Dilemma game and absent a spirit of trust, fairness, and shared rewards one or both parties easily gravitate toward

the non-cooperation/low productivity option. One of the problems with competitive markets and a "commodity" employment relationship is that they undercut all of these supports of cooperation and cause workers and managers to instead view each other as short-term, adversarial, zero-sum bargainers in which the goal is a one-off "buy low/sell high" outcome. Stabilizing employment and providing job security, on the other hand, gives workers a greater stake in cooperation (they will be around to reap the rewards), while maintaining wages, work conditions and HRM programs that create a climate of procedural and distributive justice helps build trust among workers and thus likewise promotes cooperation and efficiency. The marginal product of labor, therefore, is not a technological datum from the production function but a variable quantity influenced by a host of workplace conditions and managerial practices. Since fairness is frequently evaluated *relative* to the treatment or outcome of others, interpersonal comparisons, status rankings, and feelings of envy and jealousy (all forms of interdependent preferences) are endemic to the workplace and successfully managing these relativities becomes very important to efficiency and profit.

Say's Law and Under-Consumption: for all the reasons cited, most labor markets are not a good first approximation to the competitive model of neoclassical theory. Flexible wages cannot and do not act as an effective equilibrating mechanism in all but perhaps the very long-run; rather, labor markets often remain out of equilibrium (in the sense of a demand/supply imbalance) for months and years and restoration of equilibrium comes about as much from labor quantity and quality adjustment (e.g., demand/supply curve shifts) as from wage adjustment. At a macro level, even with completely flexible wages the aggregate labor market is not self-correcting (as maintained by Say's Law) since wage decreases reduce household income and aggregate demand (per Keynes) and, hence, cause production and employment to depart even

further from a full-employment equilibrium. Likewise, a free market economy is prone to under-consumption in the medium-to-long run because the bulk of the fruits of productivity growth are distributed to a relatively small group in the top part of the income distribution who have more inelastic labor supply curves (due to scarcity of unique talents, skills, positions) and thus reap a proportionately large part of real wage gains from economic growth (e.g., like CEOs and entertainment/sports stars). Since these groups have a lower propensity to consume, capital investment and stock market speculation start to run ahead of increases in lower/middle class income and consumer spending, leading to emergence of an excess supply imbalance and eventual downturn and even crisis – perhaps postponed for a considerable time by increased borrowing and debt.

IE Perspective on Unions

I offer two extended quotations to illustrate the IE perspective on unions, followed by a more detailed summary of key points.

The first quotation comes from Harry Millis in Congressional testimony on the then-proposed National Labor Relations (Wagner) Act. Millis was an institutionally-oriented labor economics professor at the University of Chicago and president of the American Economics Association in 1934. He states (National Labor Relations Board, 1985, pp. 1553-4):

> Of course if there were perfect mobility of labor, keen competition for labor, and no concerted control of wages and hours by employers, the situation would be substantially different from what is has been and the case for collective bargaining would be less conclusive in modern industry. I am aware that many of my academic brethren assume that these conditions just mentioned are generally true, and reason that in the absence of such friction in the market, wages, hours, and all the rest of it rather steadily adjust themselves to what industry, and consumers, should and can bear."

But then he notes (pp. 1553-54),

> The truth, as I see it, is ….. that the competitive demand for labor, while important, does not go far in protecting the workers against long hours, excessive overtime, fines, discharge, without sufficient cause, and objectionable working conditions…. One is thus driven to the conclusion that…. hours of work and conditions of work -- things which intimately concern workmen, are best decided collectively -- through legislation or through collective bargaining, and some of them are not easily subject to legislative control. This is particularly true of a reasonable degree of security of tenure. The case for collective bargaining is only less strong with respect to wages.

Millis offered this assessment in the mid-1930s when the Great Depression and massive unemployment had destroyed many people's faith in the efficacy of capitalism and competitive labor markets; hence, he may have been led by the events of the period to over-emphasize the degree to which competitive forces are defective and the corollary need for collective bargaining. Nonetheless, writing twenty years later Lloyd Reynolds offered a broadly similar evaluation. He states (Reynolds 1954: 550):

> "Trade unions, then, do not intrude into a situation in which wage rates have been perfectly aligned by competitive forces. They come into a situation in which relative wage rates have already been distorted through the failure of competition to function effectively. One cannot start from a presumption that unionism, by substituting control for competition, pulls wages away from a pre-existing ideal relationship. This ideal relationship exists only in the minds of economists. In actuality, trade unions may serve either to correct previous distortions of the wage structure or to perpetuate or accentuate these distortions. Which of these things unionism actually does is what we have to find out."

Reynolds' assessment of unionism is now more than six decades old and, like that of Millis, in part reflects conditions of the times (e.g., the predominance of manufacturing; the lingering effects of the Great Depression). With due recognition to changing conditions and diverse and evolving ideas among IE writers, it is nonetheless possible to identify what I think are relatively enduring and widely agreed-upon key points about the IE theory and perspective

on unions. These propositions to some degree apply to all unions but have most relevance to the Anglo-American type of industrial relations system. Among these are the following.

Labor Combination, Workers' Interests, and Freedom of Association: The law in non-socialist countries gives owners of capital legal permission and even encouragement to combine their resources in institutions called "corporations" for advancement of their collective interests. Both the law and economic theory find reasons why such combinations of capital promote the social interest, at least if kept within certain bounds of size and power. The question is: should the law give similar permission and encouragement to owners of labor to organize for collective action? For several centuries the law came to a negative conclusion on this matter and only reversed course in a decisive way in the USA with enactment of the NLRA and other New Deal labor legislation. After a decade or two the main current of legal enactment legal scholarship began to slowly drift back toward a more negative verdict, particularly under the influence of the Chicago-led law and economics movement. Much the same is true for economic theory, except that the shift to a more positive perspective was shallower and much less widespread among economists and more quickly swung back to the traditional negative perspective. Institutional economists were the major group in the economics profession to press for a more even-handed treatment of trade unions ("labor corporations") in the early 20th century and they continue make this case. They do so on three grounds.

The first is social justice and parallel equity. If capital can combine, and given that the interests of capital and labor are to some extent opposed, it is only fair and reasonable that labor be allowed to combine in order to preserve a balance of power and prevent exploitation and subordination of one side by the other. Second, neoclassical economic theory comes to a negative verdict on labor combination and protection because it assumes the sole end of

economic activity is maximizing consumers' welfare via a Pareto optimal resource allocation and array of final goods and services. Labor is modeled as a factor input no different from coal and computers (one of the X_i inputs in the production function) and its contribution to social welfare comes from producing the most goods at the least cost. Whether this contribution entails an 8 or 12 hour work day, a high or low wage, or adult or child labor is given zero independent weight in the social welfare calculus. From an IE perspective, however, labor is a distinctly human factor and the terms and conditions of labor should be given some explicit weight in society's welfare function. It is only common sense, after all, that people gain happiness in life not only from having more and cheaper cars and clothes but also satisfying, safe and secure jobs. Finally, progressive societies accept as a fundamental human right that among the freedoms enjoyed by their peoples is freedom of association, including freedom of association at the workplace.

Economic and Political Functions: Unions have both economic and political functions and these take place both at the workplace/industry and economy/society levels. Freeman and Medoff (1984) allude to his dual role in their "two faces" concept of unions; what they did not acknowledge is that this distinction in its generic and broad-based form goes back many decades in the IE labor literature.

The economic function of unions is to improve the terms and conditions of employment for their members by replacing individual bargaining between employer and worker with collective bargaining between the employer and workers represented by a union negotiator. Because of the strike threat and other pressure tactics, the union has more bargaining power ("muscle") than the individual worker and can thus "level the playing field" in wage determination and win higher wages and other benefits for the members. These gains spill-over to benefit employees at other companies through pattern bargaining, the union threat effect at

non-union companies, and other such methods; another form of positive spill-over is at the macroeconomic level where unions increase the wage share of national income and thus promote greater household spending, growth in aggregate demand, a buoyant job market and economy, and a "shared prosperity."

The political function of unions at the workplace is to replace a governance system of employer monarchy (the "divine right of capitalists") with a more democratic, representative, and bilateral system in which workers have a say ("voice") regarding the specification and administration of the rules they work under, provision for "due process of law" in resolving disputes over work performance and rule enforcement, and protection from arbitrary/unfair discipline and discharge. Unlike Freeman and Medoff (1984), however, IE recognizes that employee voice by itself (i.e., communication) is likely to have a slim-to-negligible effect if workers lack power to motivate employers to take their suggestions and preferences seriously. The Webbs (1897) called a bilateral system of representative employee voice "industrial democracy" and Commons (1905) called it "constitutional government in industry." At the national level, unions use their political muscle to win expanded economic and social rights for union workers and labor as a class. Particularly in early-to-middle stages of economic modernization, unions are part of a social movement that pressures governments and national elites to change the laws of the land so labor shifts from being a marginalized/disadvantaged "outsider" to a fully participating and justly treated "insider" living in a progressive welfare state where workers have economic security and opportunities for the "good life."

Unions as Organizations: Unions are political organizations operating in an economic environment. In their collective bargaining function they act as an agent for their worker-members. Although often called a "labor monopoly," unions as the agent of workers function

more like a cartel in that they negotiate with the employer(s) a uniform price schedule for labor services and then let the individual worker-members enter into employment (labor rental) contracts at this rate. Akin to other kinds of cartels, unions can successfully maintain an above-market wage, particularly over the long-term, only by controlling most of the relevant labor supply and preventing insider cheating on the agreed-upon wage schedule. Unions are formed by workers and the workers to varying degrees control the union through an electoral process. They exercise pressure and guidance through various means, such as selection/election of officers, contract ratification votes, strike votes, and quitting the union ("voting with their feet"). Unions are imperfect representative agents; some are highly democratic but others are oligarchic and "boss ridden." Due to principle-agent problems union leaders have varying degrees of discretion to pursue their own interests and agendas. Unions are not monolithic institutions, therefore, and to some degree union leaders and the individual members have different objectives.

Union Bargaining Goals: As a general statement it may be said that the bargaining goal of unions is "more" in the sense of more wages, benefits and all other workplace "goods." To some degree unions follow the path of least resistance in determining where their pressure for more is directed; whatever the exact form, the result is a redistribution of wealth and workplace control from capital (and consumers) to labor. For purposes of theory-building the economic dimension of "more" can usefully be boiled-down to two essential attributes for each union member: the *wage* (broadly defined to include all forms of compensation) and the *job*.

An IE model of unions begins with a union objective function and two principle constraints (Kaufman 2002). The objective function must satisfy two conditions: first, it is aggregated from individual preferences of the members via the union's internal electoral process and, second, it takes into account the distribution of property rights among the membership and

the resulting distribution of benefits and costs among the members from alternative wage policies. Among the alternative models so-far developed, the median voter model of unions is for analytical purposes the best-suited and most insightful. As the union seeks to maximize the preference of the median union member, the union as an organization and each individual member confront two fundamental constraints. The first is the threat of job loss, as given by the position and elasticity of the labor demand curve; the second is the threat of additional bargaining costs in the form of days out on strike, as given by a strike duration function. The higher are prospective job losses and/or strike costs, the more the union is led to moderate its bargaining demands.

Monopoly Power for Good and Bad: Regardless whether unions are envisioned as a cartel, labor monopoly, or some other institutional form, they bring market power to the seller's side of the labor market. In a non-union system the typical individual seller of labor has no power to gain a wage higher than dictated by the market (a competitive situation) or the firm (a monopsony situation); a union, on the other hand, gains market power from the strike threat and other sanctions and changes the wage for the members from a parametric "given" (take it or leave it) to a bargained and therefore adjustable rate.

This fact is non-controversial and both supporters and opponents of unions agree on it. The controversy comes from whether union power over wages and other terms and conditions of employment is a force for good or bad with respect to economic performance and social welfare. In standard neoclassical analysis, the base-line answer is "bad" because the analysis starts off with a competitive theory of labor markets and a production function theory of firms. Depending on the author, selective qualifications may be introduced here or there that modestly attenuate the "bad" answer; practically never, however, is the possibility of a "good" answer given serious

credence. An IE theory of unions, on the other hand, provides a more nuanced and balanced assessment. IE leads to two implications: first, on theoretical grounds union power may be either good or good for performance and welfare and (2) over time (i.e., over repeated bargaining rounds) the verdict probably progressively shifts toward the more negative. I postpone consideration of the second point to the next section.

Regarding the first, we get a clue from Commons' (1913) observation that the economic goals of unions are *wealth redistribution, aggrandizement and protection* (WRAP), that a union can only accomplish these goals by in some way "preventing the employer from doing as he please," but that this constraint cannot automatically be condemned because the verdict depends on "whether its [the union's] restrictions are injurious or beneficial?.... and to whom?" (p. 136). A second clue is provided by Slichter (1931) who states, "The monopoly may succeed in raising the price and even raising it substantially, but this does not necessarily mean the price is unreasonably high… The best examples of monopolies which merely eliminate cutthroat competition are found among labor unions" (p. 365). When we bring in the political function of unions (constitutional government in industry, CGI), the result is a 2 x 2 matrix of union welfare effects, such as given in Table 1 (Kaufman 2007c). Along the top are the two functions of unions: economic and political; down the left-side are two alternative characterizations of the union effect: "aggrandizement/restriction" (AR) and "protection/improvement" (PI). The idea, following Commons and Slichter, is that union market power has two alternative effects. The first is a *negative* performance/welfare effect in which union power is an aggrandizing/restricting force that shifts the wage determination process (external labor market) and governance function (internal labor market) from an approximately balanced ("competitive") situation to a lop-sided monopoly situation in favor of labor. The second is a *positive* performance/welfare effect in

which union power is used to improve substandard economic conditions in the firm/industry and protect workers from oppression and injustice, thus shifting the employment relationship to a more balanced situation and ending employer dominance.

[Insert Table 1 here]

Different theories fill in the four possible cells in different ways. The standard neoclassical model occupies cell I (union as a monopoly-like cause of resource misallocation) and probably cell II (monopoly governance by union bosses, internal electoral oligarchy, political rent-seeking) and most often either ignores or marginalizes cells III and IV. The "efficient contract" model (Kaufman 2002; Naylor 2003) also occupies cell 1 but only marginally so since the union's redistribution of wealth from capital to labor does not disturb the initial (assumed) competitive resource allocation (it might, however, negatively affect investment and future growth). It is largely silent, on the other hand, about the political function. Freeman and Medoff's (1984) "two faces" of unions theory starts-off occupying cells I and IV – unions are a negative monopoly force in wage determination but a positive "voice" force in internal firm governance and national politics. F&M argue, however, that the voice function leads to positive indirect economic effects (lower turnover, higher productivity, etc) and the net outcome for the economic function thus moves from negative to zero or even positive – thus moving modestly, perhaps, into cell III as the final outcome.

The IE perspective on unions, by way of contrast, opens up all four cells -- and all permutations therefore -- as possible outcomes. In this respect IE postulates unions have "four faces," not two. That is, since labor markets and internal firm governance and national politics can be employer dominated, union power in this case serves a PI role and acts to restore

(approximate) economic and political balance (cells III and IV). Conversely, unions may also dominate external markets and become an overly aggrandizing/despotic force in workforce governance and national politics (the AR role) and thus harm performance/welfare (cells I and II). One may telescope the "four faces" of cells I-IV into "two faces" by focusing only on the left-hand column – a positive *Protection-Improvement face* and a negative *Aggrandizement-Restriction face*; alternatively, by defining monopoly and monopsony very broadly to include both economic and governance functions one may also call the positive face of unions *monopsony-reducing* and the negative face *monopoly-creating*.

The IE position, therefore, is that economic theory provides no *a priori* judgment on the performance and welfare effects of unionism; therefore, the verdict must be considered on a case by case basis and ultimately decided on the strength of the empirical evidence. This position is well stated by Reynolds (1951): "Whether the results of collective bargaining are better or worse than those of unilateral wage administration by employers…. is basically a problem for empirical study" (p. 260).

Flaws in the Neoclassical Analysis of Unions

The neoclassical model leads on *a priori* grounds to the conclusion that unions most likely harm performance and welfare (recall the quotations featured earlier by Addison and Hirsch and Booth). IE allows that unions may indeed have a negative effect and in no way rules this out; what it also maintains, however, is that on equally solid theoretical grounds unions may have a neutral or positive effect. The most direct and obvious way to resolve these conflicting hypotheses is to "crunch the numbers" and see what the empirical evidence reveals.

Unfortunately, assembling clear-cut and convincing empirical evidence is difficult because available data sets and statistical methods often lack sufficient powers of discrimination; further, it seems fair to say that a certain element of theoretical/ideological dogmatism on the side of both union opponents and proponents tends to preclude the possibility that *any* new empirical evidence can change minds. Hence, a more indirect solution to the dispute is to move back a step and reconsider if the negative verdict of neoclassical theory is really as iron-clad and compelling as presented. This is the tact taken in this section using IE ideas and concepts.

An important reason NLE comes to a negative verdict on unions is because it starts off the analysis with a model of a perfectly competitive labor market. Since resources are already efficiently allocated and outcomes are Pareto optimal, it is asking a lot for a trade union to then improve on this situation. Indeed, economist Harold Demsetz (1969) calls this type of welfare analysis the "Nirvana fallacy" because it asks proponents of an imperfect flesh-and-blood institution (a union) to demonstrate that it can improve the performance of an economy that is already assumed perfected "as if" by an Invisible Hand guided by an omniscient auctioneer. IE reaches a different conclusion about unions because it embeds them in an alternative model of imperfect competition.

Earlier generations of IE labor economists rejected the competitive model as an appropriate tool for rendering welfare judgments on trade unionism (Taylor and Pierson 1957; Kaufman 2008). Their position regarding the competitive model was built on two lines of argument. The first is empirical evidence that numerous labor market outcomes (e.g., unemployment, non-compensating wage differentials, occupational segregation) fail to match the predictions of the theory; the second is that a number of the assumptions of the competitive model (e.g., flexible wages, few constraints on mobility, futures markets in labor) are not only

drastic abstractions (a necessary concomitant of theorizing) but *falsifications* of reality. Neither of these lines of attack proved successful, however. Neoclassical economists were able to extend price theory to explain why the alleged imperfections were themselves efficient outcomes (e.g., rigid wages stem from implicit contracts; unemployment arises from efficiency wages; job search is a rational adaptation to scarce information), thus seeming to take the sting out of these apparent anomalies. Likewise, Friedman (1953) persuaded economists that the IE position rested on the proposition that theoretical assumptions should be *descriptively* realistic – a patently silly proposition that economists naturally rejected -- when, in fact, the IE position insists on *substantive* realism (empirically congruent – or at least not falsified -- priors). Despite the straw man nature of this argument, it proved highly persuasive and economists to this day instinctively and almost *en masse* turn away from any critic of the competitive model who starts to question the model's assumptions.

Since the competitive model seems largely impervious to criticism on grounds of *external empirical* inconsistency, the only other route for critics is to look for places of *internal logical* inconsistency. IE pinpoints three sources of logical inconsistency/contradiction in the competitive model of labor markets. Since these points have been developed elsewhere (Kaufman, 2007d; 2010), I summarize for sake of brevity.

The first reason is that the model of perfect competition implicitly assumes zero transaction cost (TC), which is to say that property rights to goods and services can be exchanged at no cost. Zero TC also implies a world of complete and perfectly enforced employment contracts. As earlier described, IE theory notes that economic activity can be coordinated by (at least) two alternative means: markets using competition and price and hierarchical organizations using command and administration. As deduced by Coase (1937), the

economy will gravitate toward the coordination mode that is most efficient (least cost). If TC = 0, markets are everywhere more efficient than organizations and the economy dis-agglomerates into a perfectly decentralized system of single-person firms in which the entire division of labor is coordinated by demand and supply. However, single person firms by definition have no employees so in this world the employment relationship and the labor market (in the form of a factor market) disappear. In effect, the model of a perfectly competitive labor market used by economists to evaluate unions is a theoretical case of "too good to be true" for it is so perfect that firms do not need employees (who may shirk, show up late for work, want health insurance, etc.); instead, they can hire independent contractors (and perhaps also rent them a portion of the capital stock, thus preserving the economies of large scale production) who do the work for the exact price and specifications written in complete sales contracts. The upshot, then, is that *panel a* in Figure 1 has no theoretical grounding (i.e., it is *ad hoc* and lacks solid micro foundations) and therefore its analysis and conclusions regarding unions are equally compromised.

The second reason is the converse argument that labor markets are always and everywhere imperfect; hence, the appropriate model is not perfect competition but imperfect competition – including but not limited to monopsony (broadly defined). Two arguments point to this conclusion. The first is the converse of the argument in the previous paragraph. If labor markets and employment relationships cease to exist in a world where TC = 0, the logical implication is that the necessary condition for their existence is TC > 0. But positive TC arises from positive and possibly large "frictions," such as imperfect and asymmetric information, indivisible or missing property rights, and limited human decision-making ability. Hence, labor markets are by their nature imperfect and can never be otherwise *even in the world of theory*. A second reason why labor markets are inherently imperfect is because labor is not a homogeneous

commodity as presumed in NLE but is delivered by a heterogeneous human being who necessarily forms a personal relationship with management (rather than an anonymous market exchange). Heterogeneity makes the labor supply curve upward sloping, thus introducing a form of monopsony; likewise, the human nature of labor – in conjunction with incomplete contracts – introduces a host of contracting problems into the employment relation, such as principal-agent, moral hazard and adverse selection.

The third reason is that the model of perfect competition gives a misleading estimate of the negative performance/welfare effect of unions (such as in *panel a* in Figure 1). Two arguments support this contention. The first is that in a world of perfect competition (TC = 0) we know from the Coase theorem that efficient resource allocations are "institution neutral" and agents can through private bargaining exhaust all gains from trade. Hence, a union-induced misallocation of resources and attendant deadweight loss disappear in a world of zero TC as unions and firms are able to bargain their way to a point on the contract curve, as in the efficient bargain model of unions. Thus, one concludes that the standard NLE analysis of the welfare costs of trade unions in *panel a* is logically flawed; it predicts that in a competitive market unions cause a reduction in efficiency when in fact they cause no such loss (even though a redistribution of wealth takes place). The second argument takes the logic in the opposite direction. The union wage is point B in *panel a* and it appears that the economy suffers from a deadweight loss. The size of the triangle ABC is overstated, however, because it is comparing two situations that are in fact non-comparable. The competitive outcome at point A presumes TC = 0; by the reasoning just given, however, the union outcome at point B requires (to be durable) TC > 0. But the cases of TC = 0 and TC > 0 preclude a consistent welfare comparison; in effect the TC = 0 situation places the economy on a higher production possibility frontier than the TC >

0 situation (since with perfect information more can be produced with given land, labor and capital). Thus, in terms of *panel b* of Figure 1, the competitive market outcome is at point A and the union outcome is at point B; point B can*not* be judged as worse than point A, however, because it is impossible for the economy to actually reach point A *even if the union disappears* and the labor market returns to a state of pure (but not perfect) competition. In other words, making a welfare comparison between points A and B in *panel a* commits the Nirvana fallacy.

An IE Model of What Unions Do

The next task is to take the IE insights about unions, particularly as encapsulated in Table 1, and represent them in a more analytical manner. This is done with the help of Figure 2.

Panel a starts with a purely competitive labor market. The non-union wage is W_1 (point A). Now introduce a union into this market. The task of the union is to choose a wage rate it will demand from the employer. For simplicity, this process is divided into two steps: in step one the union canvases its members for a preferred wage rate and at this point they consider only the labor demand curve constraint; then in step two the union puts an actual wage demand on the table but at this point with consideration of the strike cost constraint.

A key part of the IE theory of unions is that they are political organizations. Thus, this consideration must be taken into account in formulating the union's objective function. As earlier described, to obtain the union objective function one must know the distribution of property rights in the union and the nature of its internal governance system. The simplest and most tractable approach is the median voter model. Assume the jobs are rationed, as in most industrial unions, by a "first in-last out" seniority rule. Given a downward sloping labor demand curve, this

rule creates a well-ordered distribution of most preferred wage demands on the part of the rank and file. Assuming perfect democracy in the union, ruling out principal-agent problems (no leadership discretion), and assuming the leadership is elected by majority vote, one can deduce that the union's optimal wage demand is the preferred wage of the median union member in the seniority distribution. This wage demand is given where the median voter's L-shaped indifference curve (I_M) is tangent to the labor demand curve (Point B).[1] If for simplicity the member's time horizon is limited to a single contract period, the union's wage demand is W_2, L_M, where L_M is an employment level where the median member remains employed but is the next to be laid-off. At this point the union has extracted the highest attainable "more" for the median member, subject to the labor demand curve constraint.

Given only the labor demand curve constraint, the union's preferred wage/employment outcome is W_2,L_M. The standard monopoly model assumes the union has the power to dictate this wage to the company; hence, the final negotiated wage is W_2 and the welfare cost of unionism can be calculated, as in Figure 1. In IE, however, positive TC makes bargaining ubiquitous, even in non-union employment relationships, and part of the positive rationale for unionism is that collective bargaining offsets the bargaining weakness of the individual worker. Hence, in an IE model the bargaining process and bargained outcome are essential parts of the story, leading to phase two of the theory. From internal caucus the union leadership determines via the median voter process that the membership's preferred wage is W_2. They now enter into negotiations with employers. The negotiation process, given imperfect information and bounded rationality, leads to a dynamic process of offer and counter-offer as the two sides slowly iterate

[1] Each member's indifference curve is vertical at the level of employment at which he/she is next to be laid-off, showing that utility falls to zero (or the value of the alternative wage) if the job is lost; to the right of this point the horizontal segment of the indifference map shifts upward at each higher wage, indicating utility rises with higher wages *as long as* the member is employed. The L shape indicates each person is willing to trade-off zero in wages to preserve someone else's employment (pure selfishness).

toward agreement. This process is represented in *panel b* by the union's "resistance function" U(W) and the firm's "concession function" F(W). This model was first presented by Hicks (1932) and variants (e.g., a two equation set of reaction functions) have been formalized and extended (e.g., Kaufman 1981; Andersen and Devereaux 1988).

Panel b has the wage on the vertical axis and time (days spent in bargaining) on the horizontal. To force closure in bargaining, the union sets a strike deadline at T_1. The prospect of strike costs leads the union to lower its initial wage preference of W_2 in *panel a* to a first-round bargaining demand of W_5 in *graph b*; likewise, the company by the same logic moves from a preferred wage of W_1 to an initial offer of W_3. As the strike deadline approaches, the two sides make further concessions until at the strike deadline they reach agreement at W_4 (point C). Although most contract negotiations are settled without a strike, in some cases the parties are so far apart or so intransigent that the two curves do not intersect at the strike deadline and a strike takes place. Assuming the relationship does not break apart, however, at some point an agreement will be reached, either at W4 or some other nearby wage.

The conventional monopoly model assumes labor markets are competitive and, further, typically omits the bargaining process and strike cost constraint.[2] Hence, a union appears to be a relatively straightforward case of labor monopoly, a source of significant wage distortion (W_2-W_1) and employment loss (L_1-L_2), and cause of resource misallocation. The institutional model leads to less negative conclusions – at least in the short-run and, indeed, suggests that unions may well be a source of efficiency gain. The "less negative" part of the conclusion comes from

[2] Kaufman and Martinez-Vasquez (1987) incorporate both the demand curve and strike cost constraints and show that the median member's preferred wage is determined by the one that binds first. This insight offers an explanation of the Ross-Dunlop debate; that is, whether unions do or do not take into account the employment effect in formulating their optimal wage policy. If the labor demand curve is the binding constraint, the union takes into account the employment effect (a "Dunlop-type" union); if strike costs are the binding constraint then it does not (a "Ross-type" union).

including the bargaining process and strike cost constraint, as in *panel b*; doing so, for example, lowers the negotiated union wage to W_4 (from W_2) and, therefore, reduces the extent of resource mis-allocation in a competitive market situation.[3]

The "efficiency gain" part of the IE perspective requires a shift of attention to *panel c* of Figure 2. IE theory leads to the conclusion that *all* labor markets are imperfect; hence, such a situation is pictured in *panel c*. The labor supply curve (S) is upward sloping, for reasons described earlier; accompanying it is a rising marginal cost of labor schedule (MCL). In imperfect competition, the firm(s) have a continuous marginal revenue product schedule (MRP) but not a well-defined labor demand curve (Fleischer and Kniesner 1980). In this situation the employer has some degree of market power and, as in any kind of monopsonistic market, is able to pay a wage such as W_6 – a wage that is *lower* than the competitive wage W_1. Employment L_6 is determined by the intersection of the MRP and MCL schedules (point E) and is less than the competitive level L_1.

The bargained wage of W_4 (*panel b*) leads to resource misallocation in a competitive labor market (*panel a*). In an imperfect labor market, however, the reverse may well be true. This is the situation pictured in *panel c*. There the union wage raises the wage W_4 closer to the wage W_1 that would exist in a competitive market; given W_4, employment in turn rises from L_6 to L_4 (from point E to F). In terms of the production possibility frontier pictured in *panel b* of Figure 1, the existence of imperfect competition in the labor market results in a misallocation of resources (such as point B) and the advent of unionism moves the economy closer to the frontier and an efficient allocation (closer to point A).

[3]The strike cost constraint lowers the bargained wage, thus reducing the negative monopoly efficiency effect, but one must also take account that bargaining costs in general and strike costs in particular use up valuable economic resources. The quantitative size of the latter appears to be minimal at an aggregate long run level, however. See Gunderson and Merlino (1987).

Consistent with the IE perspective, Figure 2 shows that unionism can either be monopoly-creating (*panel a*) or monopsony-reducing (*panel b*). The analysis in Figure 2 is explicitly framed in terms of wages but, as noted earlier, monopoly and monopsony for this discussion should be framed broadly to include both the economic (WRAP) and political (CGI) dimensions of unions. In its monopsony-reducing role, for example, unions not only offset employer power in the external labor market but also counter-balance employer domination in internal firm governance and the polity.

Extensions and Qualifications

This base-line IE theory of unions needs to be extended and qualified in several directions.

First, it is useful to present an alternative depiction of the IE theory on "what unions do." This is shown in Figure 3. The diagram shows a frequency distribution of firms with respect to the breadth, depth and tenor of their HRM/employee relations practices. The distribution may be for a particular industry or for the entire economy; the distribution depicted in Figure 3 reproduces the actual pattern found in the contemporary U.S. economy (Freeman and Rogers, 1999; Kaufman, 2010b). The horizontal axis (HRM/ER) extends from "lowest" (most primitive and oppressive type of employment system) at the origin to "highest" (most advanced and progressive type of employment system).

The competitive model of labor markets assumes homogeneous firms and workers; the former arises from assumptions such as constant returns to scale and the presumed existence of a "no economic profit" long-run equilibrium. IE, however, assumes heterogeneity among firms and workers, reflecting market and entrepreneurial attributes in a positive TC situation such as

differences in managerial ability, organizational strategies, production technologies, product and employee characteristics, training systems, and control of strategic resources. The consequence of this heterogeneity is that firms differ considerably in product and labor market performance, as measured in terms of profitability, productivity growth, customer service, and terms and conditions of employment. Because of this heterogeneity, along with less-than-perfect competitive selection pressures, some firms through deliberate choice or poor management occupy the bottom end of the HRM/employment distribution. These firms include sweat shops, low road employers, and the "100 worst companies to work for." They may not well fit the model of monopsony narrowly interpreted, but broadly interpreted they are monopsonistic in the sense of a substandard and perhaps exploitative economic package of wages, benefits and conditions and a substandard and often oppressive/unjust system of workforce governance.

It is fair to say that this "less than competitive" segment of firms is almost entirely omitted from the standard neoclassical story; also omitted is any attention to the possibility that workers may need *protection* from the exploitative or hard-pressed companies that dwell in this segment and that "protectionism" provided by unions may for this reason promote rather than harm the social welfare. From an IE perspective, the purpose of unions (and forms of protective labor legislation, like a minimum wage law) is, in part, to establish a floor of minimum acceptable terms and conditions of employment that no firm may go below (Kaufman 2010a). A union does this through collective bargaining and in *panel c* of Figure 2 the floor is set at the wage W_4; in terms of Figure 3 and HRM/employee relations practices the floor is represented by the vertical line and the union in effect cuts-off the lower end of the distribution below HRM/ER_1.

The neoclassical model suggests that when higher union labor costs curtail employment and put firms out of business this harms economic performance and welfare. IE, suggests, however, that eliminating the lower end of the employment distribution, such as in Figure 3, may well improve performance/welfare. From an IE perspective, these firms survive and perhaps even prosper by providing less than competitive terms and conditions of employment; it is wise social policy, therefore, to put them out of business – or force them to improve labor conditions – so the capital and labor resources can be used in a more efficient and progressive manner. In effect, the "low standard" firms in the bottom part of the employment distribution are "parasitic" in that feed on resources that should go to labor but which due to competitive imperfections are available for capture by employers. These firms are also often laggards in managerial practice, technological sophistication, and product quality; the effect of unionization, therefore, usefully forces them to become competitive not through further skimping on labor conditions but by improving the other areas of their business. If unionism results in a decline in employment, therefore, it may not be detrimental to performance/welfare as assumed in neoclassical theory; rather, it may be the result of forcing firms to shoulder the full social cost of labor.

From an IE perspective, the proper role of unionism (along with protective labor law) is to police the bottom-to-lower middle end of the labor market and use the threat of organization to maintain and improve standards among the remainder of non-union firms. Unionism, on the other hand, is from a social welfare viewpoint typically not needed -- and often not desired by the majority of workers – in the middle-to-upper end of the employment distribution. These firms provide wages, benefits, conditions and treatment that are range from competitive to far above average and the restrictive practices and climate of adversarialism that often go with unions on balance retard rather than promote performance and welfare. Commons (1921) states

the IE position thusly: “Labor has not come into existence at all to deal with that first class of employers [firms in the middle-upper end of the distribution]. It has come in solely in order to use coercion with …. those who need it because they will not or cannot meet new conditions” (p. 15).

Public policy through laws and regulation help determine the ease with which union can organize new firms and exert their bargaining power over wages and conditions. With this in mind, some economists (e.g., Freeman and Medoff 1984) have endeavored to determine the “optimal” level of union density, with the implication that public policy should be structured to (more or less) yield this value. Some IE writers (e.g., Dunlop) reject this way of thinking and argue on philosophical grounds that the optimal level of density is whatever workers choose – as long as it is uncoerced and determined through democratic procedures (Kaufman, 2002b: 336). A different position, however, holds that efficiency is part of an IE-specified welfare function (along with equity and human development) and thus must be given due weight in policy formulation (Budd 2004; Kaufman 2005). The model developed here suggests that the optimal level of union density is the unionization of all the firms in the shaded part of the HRM/ER frequency distribution; further, it suggest that the more imperfect and lop-sided are labor markets and firm governance the higher is the optimal union density (ceteris paribus).

Even at the conceptual level four problems arise with implementation of this idea. Number one is that the optimal level of union density is partly of a function of the health of the macroeconomy. The greatest drag on labor standards is widespread unemployment and the more severe and long-lasting is unemployment the more severe are the forces of unequal bargaining power and cutthroat competition. Thus, in a depression labor market the entire distribution of firms shifts leftward and the optimal union density rises, perhaps requiring a liberalization of the

nation's labor laws. One can interpret the events of the 1930s, including passage of the NLRA, as fitting this story. Conversely, if the macroeconomy operates at full employment in most years then the entire distribution shifts rightward (more firms are pressured by competition and labor shortages to upgrade standards) and optimal union density falls. The scenario may also well fit the post-WWII period and explain why labor law and regulation slowly shifted toward a less union-friendly position and, further, why legislative efforts to enact union-friendly "labor law reform" have failed for several decades. The model also rationalizes why unions were seen as a social "good" in the depression decade of the 1930s for their economic recovery and industrial democracy roles (boosting wages and aggregate demand, protecting downtrodden workers) but then in the post-WWII period shifted in the public mind to a social "detriment" to the extent they were perceived as hindering supply-side growth, fostering inflation, cartelizing competitive markets, and protecting "deadwood" employees in firms.

Reason number two is that it is difficult to determine what is a "competitive" level of wages and labor conditions – *assuming* this is the appropriate benchmark for analysis. IE insists that the appropriate definition of "competitive" is relative in two respects: first, it must mean wages and conditions that would exist if the economy were at full employment and, second, it must mean wages and condition that reflect a reasonable balance in rights and resources on the sides of both capital and labor. Neoclassical economists are prone to draw a demand/supply diagram and infer optimal welfare properties to the equilibrium wage without attention to either of these qualifications. Further, many economists (such as from the Chicago School, see Reder 1982) then infer that existing wages and employment are approximately competitive and, hence, unions necessarily intrude into real world markets with deleterious effects. IE does not accept

this kind of reasoning because it is biased toward the status quo and ignores the fact that not all "competitive" outcomes (12 hour work days? child labor?) are in the social interest.

The third caveat further complicates the matter because IE does not accept the near-universal proposition that a competitive labor market (even if a solid logical proposition) yields optimal economic efficiency. In the real world, hyper-competitive labor markets subject workers to excessive stress and insecurity, leading to impaired decision-making (because environmental turbulence exceeds the brain's coping capacity), lower productivity (e.g., because of quick loss on all forms of investment specific to occupations, training and firms), and efforts to erect protective barriers (e.g., unions). Further, perhaps in the world of theory -- where the problems just described are banished by the power of assumption -- a perfectly competitive market yields static efficiency, yet even in theory a perfectly competitive market may not yield *dynamic* efficiency. Due to its perfect information assumption and zero economic profit equilibrium condition, the neoclassical model forecloses the dynamic forces of growth in capitalism, such as innovation, entrepreneurship and risk-taking. Hence, from an IE perspective perfect competition is not the social ideal; rather, some form of high road imperfect competition is. A modern example in the labor market is a *high performance work system* (HPWS). Empirical evidence is mixed about whether unionism on balance promotes or retards the performance of an HPWS, so this may perhaps be best left as an empirical issue to be determined (Hirsch 2007). The upshot, however, is that the optimal level of union density is probably not best pegged with respect to some "competitive" level.

Reason number four has to do with the dynamic path of union gains from collective bargaining. The neoclassical monopoly model is in one respect favorable to unions. As argued by Friedman (1951), labor monopoly is responsible for high wages but not rising wages. This

proposition presumes a static monopoly equilibrium, such as point B in Figure 1. The IE model of unions, however, suggests that in fact the union quest for "more" leads to a gradual upward creep in union wages, other direct and indirect forms of labor cost, and in restrictive or burdensome internal governance practices. This proposition comes directly from the median voter model; that is, in period 1 the union bargains for the wage preferred by the person in the median position in the seniority distribution but in period 2 this person is at the low end of the seniority distribution and the new median member wants a yet higher wage. This process repeats itself and the union slowly cuts off bottom segments of the seniority distribution as it slowly nibbles its way up the firm's labor demand curve. A more realistic specification of the median voter model where, instead of a one contract time horizon, the median member calculates the preferred wage over a multi-contract period slows the upward ratchet in wages but does not stop it (Burda 1990). Thus, over time unions transition from monopsony-reducing to monopoly-creating and lead to the oft-heard "man on the street" opinion that "unions were necessary when they started but now have gone too far." Where employers were initially parasitic the union now takes their place and feeds on capital and consumers.

This dynamic pattern in collective bargaining poses a huge dilemma for public policy because without unions the playing field is on balance tilted in the favor of companies but with unions the playing field – at least in the union sector – starts at some point to progressively tip against employers, consumers and the public interest. It also, of course, poses a huge dilemma for unions because they are in a position of slowly bargaining their way out of existence – absent new organizing. In the opinion of Slichter (1941: 366), union bargaining power typically exceeds union organizing power and hence unions slowly recede – until either some large exogenous shock (e.g., a war, depression) causes large labor discontent and a spike in union-joining (as in

the 1930s) or the government rescues the labor movement through (say) trade protectionism or expansive fiscal/monetary policy (Freeman 1998; Kaufman 2001. Whatever the case, the upshot is that the optimal level of union density is not stationary because of the dynamics of the union quest for "more."

Unions and Labor Policy

The title given to early institutional courses in labor economics was "Labor Problems." The title reflects the IE perspective that laissez-faire labor markets are inherently imperfect along a variety of dimensions and hence generate numerous outcomes that are harmful to both efficiency and social welfare. A large portion of these textbooks, therefore, was devoted to institutional solutions to labor problems. One solution was trade unions and collective bargaining; also included were progressive personnel/human resource management (at that time often labeled "industrial relations" or "industrial relations management") and government protective labor law and social insurance. Although not typically included in these textbooks, a fourth solution that Commons and others gave emphasis to was counter-cyclical full employment fiscal and monetary policies (Kaufman 2003).

If a laissez-faire labor market and macroeconomy operate inside the production possibility frontier, such as point B in Figure 1, the question is then whether some human action can potentially improve the situation. The presumption of neoclassical theory, certainly in its core Walrasian mode, is that the reason the economy operates inside the frontier is due to human-made institutional "frictions" that interfere with the operation of demand and supply and the invisible hand. Unions, minimum wage laws, taxes and rent control laws are favorite examples. Neoclassical economics, therefore, is a "non-institutional" kind of economics because

in its theory mode all institutions but the market either disappear or have no effect on production and resource allocation (per the Coase theorem and TC = 0) and in its policy/welfare mode institutions introduce imperfections and wedges that cause inefficiency and labor problems where none otherwise exist (per the first welfare theorem). The perspective of IE is just the opposite: because TC > 0 markets start out imperfect and institutions are available – if correctly structured and not overdone – to help make the economy work more efficiently and fairly. The former looks at the economy as a perfected self-regulating mechanism and deplores when imperfect humans interfere; the latter looks at the economy as a human-made device with human-embodied imperfections that can through rational planning and wise collective action be improved through institutional re-engineering.

One solution to labor problems and poor economic performance, therefore, is a socially engineered institution called a trade union. As demonstrated in the previous section, an economist can construct a theoretically-sound explanation why a union may perform a welfare-enhancing role. An IE perspective does not let the matter rest here, however. For any given economic problem more than one alternative solution usually exists and the challenge is to identify which one is Pareto superior. In this spirit Coase (1960:18) argues, "The problem is one of choosing the appropriate social arrangements for dealing with harmful effects" and Commons (1919: 185) observes, “the problem is one, not of ideals, but alternatives."

Three considerations are important for making a wise policy choice. The first is to identify the source and nature of the market/firm failure; the second is to identify the set of alternative solutions; and the third is to determine the most efficacious one.

The book *What Do Unions Do?* by Freeman and Medoff (F&M) offers a good example of this process in action; it also provides an interesting contrast between their theory of unions and the IE theory described here.

The problem F&M identify is that in non-union firms efficiency suffers because of a free-rider/public goods problem in internal firm governance that leads to an undersupply of workplace "goods" (or oversupply of workplace "bads"), such as safety (accidents), employee involvement (shirking), and job tenure (turnover). The proffered solution is more collective voice in the workplace and the alternative delivery methods considered are company-created representation councils (e.g., "company unions"), more laws, and trade unions. The solution opted for is more unionism because it offers the strongest and most durable vehicle for collective voice at the workplace.

An IE perspective suggests a considerably broader and in some respects quite different assessment (Kaufman 2007c). F&M assume that nonunion labor markets are competitive and hence the monopoly face of unionism leads to distorted relative prices and resource misallocation. IE does not accept this assumption because in a world of positive TC labor markets and firm governance are always and everywhere imperfectly competitive and to some degree employer dominated. Both neoclassical theory and F&M, therefore, omit IBP as a source of labor problems. Thus, while F&M and neoclassical theory seek to weaken union bargaining power in external labor markets (ELMs), and in addition F&M propose to strengthen collective communication inside the firm, IE theory asserts that the initial imbalance of power is likely to favor employers and that more "muscle" (power) is therefore necessary on the labor side in laissez-faire ELMs and ILMs. Indeed, without more muscle F&M do not explain how workers

are able to induce firms to implement their collective preferences expressed through voice nor to obtain the expanded representation and due process they desire.

In IE, therefore, the first issue is how far nonunion wages and labor conditions diverge from a reasonably balanced ("competitive") outcome. If outcomes for labor are in some respect below this standard, then search must commence to determine the cause. If it is a power inequality (e.g., monopsony, authoritarian "hire and fire" employers), then workers need more collective muscle. This may be in the form of greater unionism or greater/strengthened laws, the latter either in the form of enhanced rights and resources (e.g., just-cause termination, universal health insurance) or direct regulation of wages and workplace standards (e.g. a minimum wage law, workplace safety standards). If the divergence is traceable to other types of market/firm failure, however, then a different solution may be called for that leaves the balance of power relatively undisturbed but seeks to correct the cause of some other type of imperfection. In the case of F&M, for example, if unions (allegedly) have too much power over wages but there is a public goods/collective voice problem inside the firm then their proffered solution – more unions – is ill-matched to the problem (it solves the voice shortfall but at the cost of exacerbating the monopoly wage problem). The better solution would be some kind of non-bargaining/non-adversarial/non-union employee representation council, perhaps along the lines of a European works council, a Canadian joint industrial council, or American pre-New Deal employee representation plan, that provides more voice but without monopoly power over wages (Kaufman 2007c). Similarly, if the labor problem arises from asymmetric information or a negative externality then the best solution might not be more bargaining muscle (unions) but a change in law, such as a requirement that all companies spell-out workplace rules and rights in

written and legally enforceable employee handbooks or that companies give a thirty day advance notice of plant closings.

Thus, although institutional economists believe that *all* laissez-faire labor markets are imperfect and contain labor problems, they do not necessarily counsel that "more unions" are always or even often the best solution. Indeed, the opinion of Commons and many of his colleagues gradually shifted over the years from giving relative priority to collective bargaining over legal enactment as a solution to labor problems to the reverse (Kaufman 2003) – perhaps best exemplified by the remark of FDR's labor secretary Frances Perkins: "I would rather pass a law than organize a union" (Wandersee, 1993). In every situation one must weigh how well unions solve the problem relative to other institutional solutions; further, a careful weighing is required of the costs of market failure vis-à-vis "union failure" and vis-à-vis other types of institutional failure (e.g., "government failure"). Unions bring a variety of potential benefits to imperfect firms and markets but they also bring numerous costs that deserve full consideration, including the costs of strikes, restrictive workplace practices, adversarial labor-management relations, limited coverage in the market, and a secular growth in ELM/ILM monopoly power. Indeed, IE proponents of all generations recognize that unions can be relatively blunt and expensive instruments for correcting labor problems; on the other hand, they also recognize that unions are an essential component of every capitalist-style industrial relations system if balance and fairness are to be maintained in ELMs, ILMs, and the national polity (Commons 1921; Dunlop in Kaufman 2002).

The bottom line for IE proponents is that labor problems are an inherent part of employment relationships and it serves little constructive purpose to ignore, minimize or rationalize them away with "perfect world" theories; instead, economist need to search for the

most efficient and equitable solution to these problems. In this search both Commons and Coase advocate as a first line of attack that labor problems should where ever possible be solved in a way that favors voluntary market/management solutions over union or government coercion. On this matter Commons (1934) states, for example:

> "If the profit-motive…. can be enlisted in the program of social welfare, then a dynamic factor, more constructive than all others, is enlisted. It is an appeal to the business man to get rich by making others rich, and if he does not respond, then to appeal to collective action" (p. 875).

Toward this end, he advocated legal reforms to change property rights and incentives (e.g., experience rating in workmen's compensation and unemployment insurance programs); likewise, he argued the single most important cure of labor problems is government use of fiscal and monetary policy to maintain full employment – on the premise that competitive pressures incent employers to "self-cure" labor problems by practicing progressive HRM and high road employee relations. Even at full employment, however, a segment of employers remains who cannot or will not meet reasonable labor standards (Figure 3) and for these trade unionism and labor law are required tools in the fight for greater efficiency and fairness at work. The case for unions then becomes stronger to the degree that the economy mostly operates below full employment and lacks a well-developed social safety net of protective labor laws and social insurance programs.

All of this is a difficult task of institutional evaluation and design, particularly in a highly contested political environment, but as Commons (1919) remarked: "The equilibrium of democracy may not be easy to work out, but what else is there to do?" (p. 43).

Conclusion

This article has provided an overview of the institutional economics perspective on trade unions, with particular attention to a more theoretical and analytic treatment of this matter. Areas of

theoretical overlap exist with regard to the standard neoclassical treatment of unions yet, broadly viewed, the institutional theory of unions gives a distinctly different and, I would say, more balanced and inclusive picture of "what unions do."

Unions unquestionably exert market power in labor markets and internal firm governance but whether this works for good or ill depends critically on whether wage determination and governance are problem-prone, oppressive and unbalanced or reasonably efficient and fair. Neoclassical theory gives a biased answer to this question because it starts off the analysis of unions by *assuming* markets and firms are already efficient and fair (non-exploitative) via competition and the Invisible Hand when, from an institutional perspective, this matter is not a "given" but the key issue that must be determined through empirical investigation.

Even when frictions, market failures and governance autocracy are a clear source of labor problems, however, further investigation and analysis is required to determine the best option for institutional reengineering. From an IE perspective, making the profit motive work more effectively is a first place to look, since it promotes voluntarism and decentralized market solutions, but where this fails then collective action is required. The preference in this regard between expanded collective bargaining and labor law and social insurance programs must be determined by a weighing of the relative benefits and costs of each institution. Although the "optimal" level of union density is thus an empirical matter, what is certain on theoretical grounds is that unions are an essential and enduing component of a capitalist industrial relations system in order to police the bottom-middle part of the labor market, put pressure on non-union firms to maintain reasonable labor standards, and offset employers' power both in internal firm governance and the nation's polity.

References

Amable, Bruno. 2003. *The Diversity of Modern Capitalism*. Oxford: Oxford University Press.

Anderson, Simon, and Michael Devereaux. 1988. "Trade Unions and the Choice of Capital Stock." *Scandinavian Journal of Economics* 90 (1): 27-44.

Barker, George. 1997. *An Economic Analysis of Trade Unions and Common Law*. Brookfield, VT: Ashgate.

Bennett, James, and Bruce Kaufman. 2007. *What Do Unions Do? A Twenty Year Perspective*. New Brunswick, NJ: Transaction.

Booth, Alison. 1995. *The Economics of Trade Unions*. New York: Cambridge University Press.

Budd, John. 2004. *Employment With a Human Face: Balancing Efficiency, Equity and Voice*. Ithaca, NY: Cornell University Press.

Burda, Michael. 1990. “Membership, Seniority, and Wage Setting in Democratic Labor Unions.” *Economica* 57 (November): 455-66.

Coase, Ronald. 1937. “The Nature of the Firm.” *Economica* 4 (November): 386-405.

_____. 1960. "The Problem of Social Cost." *Journal of Law & Economics* 3 (1), pp. 1–44.

Commons, John. 1905. *Trade Unionism and Labor Problems*. Boston: Ginn.

_____. 1913. *Labor and Administration*. New York: Macmillan.

_____. 1919. *Industrial Goodwill*. New York: McGraw-Hill.

_____. 1920. “Management and Unionism.” In *Proceedings of the Industrial Relations Association of America,* pp. 125-30. Chicago: IIRA.

____. 1934. *Institutional Economics: Its Place in Political Economy.* New York: Macmillan.

Crouch, Colin. 1993. *Industrial Relations and European State Traditions*. Oxford: Oxford University Press.

Dow, Gregory. 1997. "The New Institutional Economics and Employment Regulation." In *Government Regulation of the Employment Relationship,* edited by Bruce Kaufman, pp. 57–90. Madison, Wis.: Industrial Relations Research Association.

Dunlop, John. 1944. *Wage Determination Under Trade Unions*. New York: Macmillan.

Ely, Richard. 1886. *The Labor Movement in America*. New York: Crowell.

Farber, Henry. 1986. “The Analysis of Union Behavior.” In *Handbook of Labor Economics,* Vol. 1, edited by Orley Ashenfelter and Richard Layard, pp. 1039-89. New York: North-Holland.

Fleischer, Belton, and Thomas Kniesner. 1980. Labor Economics: Theory, Evidence and Policy, 2nd ed. Engle Wood Cliffs, NJ: Prentice-Hall.

Freeman, Richard. 1984. “The Structure of Labor Markets: A Book Review Three Decades Later.” In *Comparative Economic Perspectives*, edited by Gustav Ranis, et. al., pp. 201-26.

_____. 1988. “Does the New Generation of Labor Economists Know More Than the Old Generation? In *How Labor Markets Work*, edited by Bruce Kaufman, pp. 205-32. Lexington, MA: Lexington Books.

_____. 1998. “Spurts in Union Growth: Defining Moments and Social Processes.” In *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century*, edited by Michael Bordo, et. al., pp. 265-96. Chicago” University of Chicago Press.

_____, and James Medoff. 1984. *What Do Unions Do?* New York: Basic Books.

_____, and Joel Rogers. 1999. *What Do Workers Want?* Ithaca, NY: ILR Press.

Friedman, Milton. 1951.”Some Comments on the Significance of Labor Unions for Economic Policy.” In *The Impact of the Union*, edited by David McCord Wright, pp. 204-34. New York: Harcourt-Brace.

_____. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.

Gunderson, Morley, and Angelo Merlino. 1987. “Estimating Strike Effects in a General Model of Prices and Quantities.” *Journal of Labor Economics* 5 (1): 1-19.

Hirsch, Barry. 2007. “What Do Unions Do for Economic Performance?” In *What Do Unions Do? A Twenty Year Perspective*, edited by James Bennett and Bruce Kaufman, pp. 193-237.

_____, and John Addison. 1986. *The Economic Analysis of Unions*. Boston: Allen & Unwin.

Hicks, John. 1932. *The Theory of Wages*. New York: Macmillan.

Kaufman, Bruce. 1981. “Bargaining Theory, Inflation, and Cyclical Strike Activity in American Manufacturing.” *Industrial and Labor Relations Review*, 34 (3): 333-55.

_____. 1988. *How Labor Markets Work: Reflections on Theory and Practice by John Dunlop, Clark Kerr, Richard Lester, and Lloyd Reynolds*. Lexington, MA: Lexington Books.

_____. 2002. “Reflections on Six Decades in Industrial Relations: An Interview with John Dunlop.” *Industrial and Labor Relations Review* 55 (2): 324-48.

_____. 2002. “Models of Union Wage Determination: What Have We Learned Since Dunlop and Ross? *Industrial Relations* 41 (1): 110-58.

_____. 2003. “John R. Commons and the Wisconsin School of Industrial Relations Strategy and Policy.” *Industrial and Labor Relations Review* 57 (1): 3-30.

_____. 2005 “The Social Welfare Objectives and Ethical Principles of Industrial Relations.” In *The Ethics of Human Resources and Industrial Relations*, edited by John Budd and James Scoville, pp. 23-60. Ithaca: Cornell University Press.

_____. “Labor Institutionalism and Industrial Relations: A Century of Boom and Bust.” *Labor History* 47 (3): 295-318.

_____. “What Do Unions Do? Insights from Economic Theory.” In *What Do Unions Do? A Twenty Year Perspective*, edited by James Bennett and Bruce Kaufman, pp. 1-45.

_____. “The Early Institutionalists on Trade Unions and Labor Policy.” In *What Do Unions Do? A Twenty Year Perspective*, edited by James Bennett and Bruce Kaufman, pp. 46-78.

_____. 2007c. “What Do Unions Do? Evaluation and Commentary.” In *What Do Unions Do? A Twenty Year Perspective*, edited by James Bennett and Bruce Kaufman, pp.520-62.

_____. "The Impossibility of a Perfectly Competitive Labor Market." *Cambridge Journal of Economics* Vol. 31(5): 775–88.

_____. 2010a. “Institutional Economics and the Minimum Wage: Broadening the Theoretical and Policy Debate.” *Industrial and Labor Relations Review* 46 (3): pp. 427-53.

_____. 2010b. “SHRM Theory in the Post-Huselid Era: Why it is Fundamentally Mis-Specified.” *Industrial Relations* 49 (2): 286-313.

_____, and Jorge Martinez-Vasquez. 1987. “The Ross-Dunlop Debate and Union Wage Concessions: A Median Voter Analysis," *Journal of Labor Research* 8 (3): 291-306.

Kerr, Clark. 1954. “The Balkanization of Labor Markets.” In *Labor Mobility and Economic Opportunity*, edited by E. Wight Bakke, pp. 92-110.

Martin, Donald. 1980. *An Ownership Theory of the Trade Union*. Berkeley, CA: University of California Press.

Metcalf, David. 2005. “Trade Unions: Resurgence or Perdition? An Economic Analysis.” In *Trade Unions: Resurgence or Perdition?*, edited by David Metcalf and Sue Fernie, pp. 83-117. London: Routledge.

National Labor Relations Board. 1985. *Legislative History of the National Labor Relations Act*, Vols. 1 and 2, pp. Washington, DC: Government Printing Office.

Naylor, Robin. 2003. “Economic Models of Trade Unions.” In *International Handbook of Trade Unions*, edited by John Addison and Claus Schnabel, pp. 44-85. Northampton, MA: Edward Elgar.

Oswald, Andrew. 1982. “The Economic Theory of Trade Unions: An Introductory Survey.” *Scandinavian Journal of Economics* 87 (2): 167-93.

Pencavel, John. 1991. *Labor Markets and Trade Unionism*. Cambridge: Blackwell.

Reder, Melvin. 1982. "Chicago Economics, Permanence, and Change." *Journal of Economic Literature* 20 (1): 1–38.

Reynolds, Lloyd. 1951. The Structure of Labor Markets. New York: Harper.

_____. 1954. *Labor Economics and Labor Relations*, 2nd ed. New York: Prentice-Hall.

_____. “Labor Economics Then and Now. In *How Labor Markets Work*, edited by Bruce Kaufman, pp. 117-44. Lexington, MA: Lexington Books.

Ross, Arthur. 1948. *Trade Union Wage Policy*. Berkeley, CA: University of California Press.

Slichter, Sumner. 1931. Modern Economic Society, 2nd ed. New York: Holt.

_____. 1941. *Union Policies and Industrial Management*. Washington: Brookings Institution.

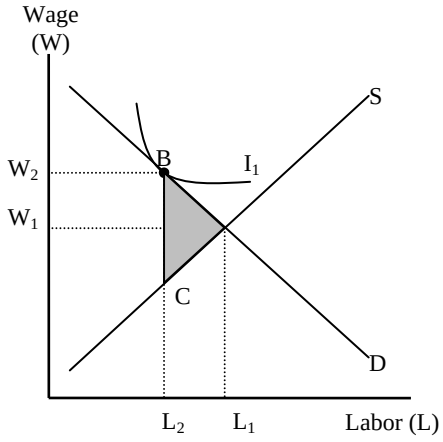
Streeck, Wolfgang, and Kozo Yamamura. 2001. *The Origins of Nonliberal Capitalism*. Ithaca, NY: Cornell University Press.

Wandersee, Winifred. 1993. "I'd Rather Pass a Law Than Organize a Union: Frances Perkins and the Reformist Approach to Organized Labor." *Labor History* 34 (1): 5-32.

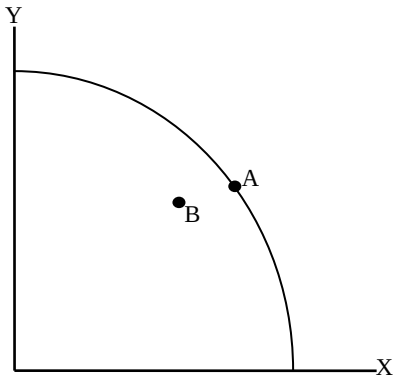
Webb, Sidney, and Beatrice Webb. 1897. *Industrial Democracy*. London: Longmans, Green.

Williamson, Oliver. 1985. *The Economic Institutions of Capitalism*. Glencoe, IL: Free Press.

FIGURE 1
Monopoly Model of Unions



Panel a: Labor Market

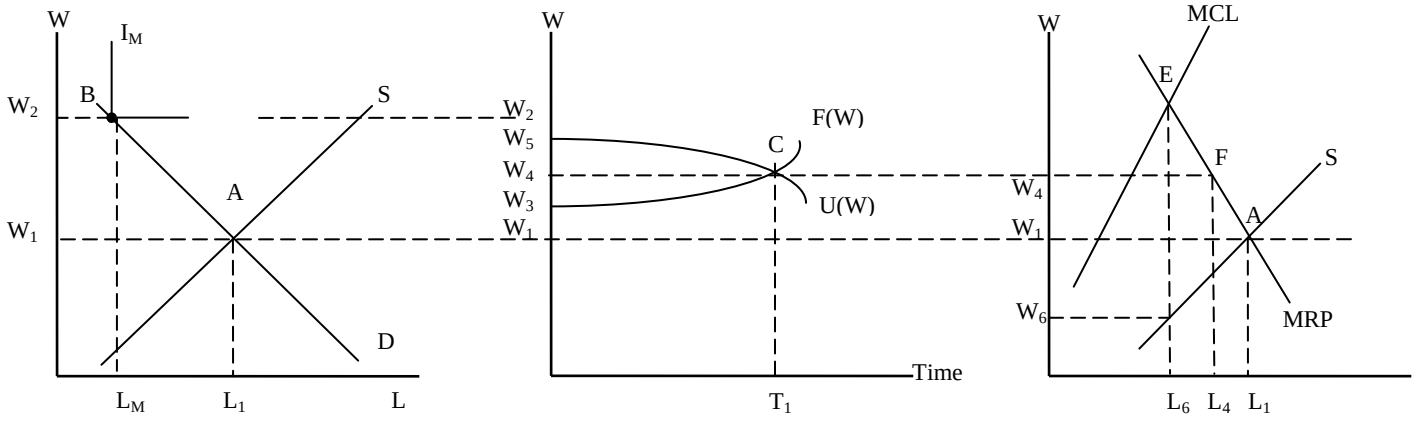


Panel b: Production Possibility Frontier

Table 1
The Four Faces of Unions:
Economic and Political Effects

	Economic Function	Political Function
Aggrandizement/ Restriction	I	II
Protection/ Improvement	III	IV

FIGURE 2
IE Model of Unions



Panel a: Competitive Labor Market

Panel b: Bargaining Process

Panel c: Imperfectly Competitive Labor Market

FIGURE 3
Optimal Union Density among Firms by Distribution of HRM and Employee Relations Practices